

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor: : **Toshihide HARA, et al.**
Filed : **Concurrently herewith**
For : **CONGESTION CONTROLLER FOR...**
Serial No. : **Concurrently herewith**

August 28, 2003

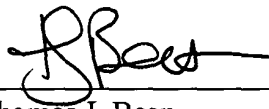
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PRIORITY CLAIM AND
SUBMISSION OF PRIORITY DOCUMENT

S I R:

Applicant hereby claims priority under 35 USC 119 from **Japanese** patent application number **2002-263102** filed **September 9, 2002**, a copy of which is enclosed.

Respectfully submitted,



Thomas J. Bean
Reg. No. 44,528

Katten Muchin Zavis Rosenman
575 Madison Avenue
New York, NY 10022-2585
(212) 940-8800
Docket No.: FUJA 20.616

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 9月 9日

出 願 番 号

Application Number:

特願2002-263102

[ST.10/C]:

[JP2002-263102]

出 願 人

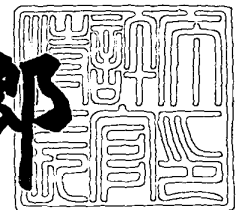
Applicant(s):

富士通株式会社

2003年 1月28日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3002203

【書類名】 特許願

【整理番号】 0250747

【提出日】 平成14年 9月 9日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 H04L 12/28
H04Q 3/00

【発明の名称】 イーサネットスイッチにおける輻輳制御装置

【請求項の数】 5

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通
株式会社内

 【氏名】 原 俊英

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通
株式会社内

 【氏名】 前田 孝英

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100077517

 【弁理士】

 【氏名又は名称】 石田 敬

 【電話番号】 03-5470-1900

【選任した代理人】

 【識別番号】 100092624

 【弁理士】

 【氏名又は名称】 鶴田 準一

【選任した代理人】

【識別番号】 100100871

【弁理士】

【氏名又は名称】 土屋 繁

【選任した代理人】

【識別番号】 100082898

【弁理士】

【氏名又は名称】 西山 雅也

【選任した代理人】

【識別番号】 100081330

【弁理士】

【氏名又は名称】 樋口 外治

【手数料の表示】

【予納台帳番号】 036135

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9905449

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 イーサネットスイッチにおける輻輳制御装置

【特許請求の範囲】

【請求項 1】 イーサネットスイッチにおける輻輳制御装置であって、
互いに異なる優先度を有する複数の送信キューと、
P A U S E フレームを受信する受信手段と、
受信した P A U S E フレームにより前記送信キューからの送信トラフィックを
抑制する抑制手段と、を有し、

前記抑制手段は、P A U S E タイム中以外の時に受信した P A U S E フレーム
によって、最も優先度の低い送信キューからの送信トラフィックを抑制し、P A
U S E タイム中に受信した P A U S E フレームによって、より優先度の高い送信
キューからの送信トラフィックをさらに抑制する、ことを特徴とする輻輳制御装
置。

【請求項 2】 イーサネットスイッチにおける輻輳制御装置であって、
送信キューと、
P A U S E フレームを受信する受信手段と、
受信した P A U S E フレームにより前記送信キューからの送信トラフィックを
シェーピングするシェーピング手段と、を有し、

前記シェーピング手段は、受信した P A U S E フレームによって、前記送信キ
ューからの送信トラフィックを予め定められたシェーピング値に基づく送信速度
以下に抑制する、ことを特徴とする輻輳制御装置。

【請求項 3】 前記送信速度の抑制は、前記送信トラフィックにギャップを
挿入することによって行なう、請求項 2 記載の輻輳制御装置。

【請求項 4】 イーサネットスイッチにおける輻輳制御装置であって、
送信キューと、
前記送信キューに滞留するパケットをその入力ポート対応にカウントすること
で、輻輳原因となる入力ポートを特定する特定手段と、
特定した入力ポートに結合される他のスイッチに P A U S E フレームを送信す
る送信手段と、を有することを特徴とする輻輳制御装置。

【請求項 5】 前記特定手段は、特定した入力ポートのアプリケーションを前記パケットの属性によってさらに特定し、

前記送信手段は、特定したアプリケーションを送信する P A U S E フレームによって通知する、請求項 3 記載の輻輳制御装置。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は輻輳制御装置に関し、特にイーサネットスイッチが複数接続される環境において輻輳が発生した際に、P A U S E フレームによって送信トラフィックを抑制する輻輳制御方式に関するものである。

【 0 0 0 2 】

【従来の技術】

従来よりイーサネット（登録商標）において輻輳が発生すると、輻輳が発生しているスイッチングハブ等のスイッチからその前段のスイッチに対して、I E E E 8 0 2 . 3 で規定される P A U S E フレームを送出して入力トラフィックを抑制している（例えば、非特許文献 1 参照）。

図 1 には P A U S E フレームによる輻輳制御の一例を、そして図 2 には P A U S E フレームのフォーマットを示している。

【 0 0 0 3 】

図 1 において、スイッチ（S W 1）1 の送信ポート側のリンク α で輻輳が発生すると、スイッチ 1 内部のバッファ（キュー）の閾値によってその輻輳を検出する。次に、スイッチ 1 は前段のスイッチ（S W 2 ~ 5）2 ~ 5 からの送信トラフィックを抑制させるため、スイッチ 2 ~ 5 へ P A U S E フレームを送信する。図 2 に示すように、P A U S E フレームは、その宛先アドレス（D A : Destination Address）に P A U S E フレームであることを示すマルチキャストアドレスを設定する。

【 0 0 0 4 】

また、そのパラメータフィールドには、例えば X X (m s) 等の P A U S E タイム (Pause time) のタイマ値を設定する。タイマ値がゼロの場合は、P A U S

E状態が解除されて送信状態に復帰することを示す。その他の各フィールドは一般的なイーサネットフレームのものと同様であり、ここではそれらについて説明しない。

【0005】

一方、受信側の各前段のスイッチ2～5は、受信したPAUSEフレーム中に含まれるタイマ値に基づいて、輻輳が発生した次段のスイッチ1（PAUSEフレームの送信元）へのトラフィック転送を一定時間抑制する。また、PAUSEフレームを受信したノードが端末の場合にも、その端末がPAUSEフレームをサポートするNIC（Network Interface Card）を有していれば、そのタイマ値に基づいて送信を一時的に抑制する。PAUSEタイムが満了すると、その間中断していた送信が再開される。

【0006】

【非特許文献1】

岡部泰一、“詳説TCP/IPプロトコル 第9回 イーサネット（その4）フロー制御とVLAN、トラブルシューティング 1. イーサネットのフロー制御”、[online] 2001年10月2日、ネットワーク技術解説講座、[平成14年9月9日検索]、インターネット
 <URL :<http://www.atmarkit.co.jp/fwin2k/network/tcpip009/tcpip01.html>>

【0007】

【発明が解決しようとする課題】

上述したように、IEEE802.3で規定されるPAUSEフレームは、輻輳を検出したスイッチ1から前段のスイッチ2～5やその配下の端末に対して全てのトラフィックの送出を一定時間抑制させるように依頼するものである。しかしながら、このような従来の動作には以下に示すような問題点があった。

【0008】

(1) QoS (Quality of Service)の確保

PAUSEフレームはトラフィックの種別を区別していないため、例えば遅延時間やジッタ条件の厳しい音声や動画像などのトラフィックと、これらの条件が

厳しくないFTP (File Transfer Protocol) などのデータトラフィックとが同等に扱われ、特に前者のQoSが確保されないという問題があった。

【0009】

(2) 実効スループットの向上

また、スイッチ1から前段のスイッチ2～5やその配下の端末に対して全トラフィックの送出を一律に一定時間抑制するため、同時期にPAUSEタイムが満了して輻輳状態と非輻輳状態との繰り返し状況が発生する恐れがあった。スループットを向上させるには、トラフィック量が波打たない定常的な状態を維持するのが好ましく、従来技術の輻輳制御では実効スループットの向上が図れないという問題があった。

【0010】

(3) トラフィックの公平性

さらに、輻輳の原因となっているフローの特定を行わずに全てのフローを止めてしまうため、存在する複数のフローに対して公平な帯域の割り当てが行われないう問題があった。すなわち、輻輳を引き起こしているトラフィック以外のトラフィックまでもが一律にその送信を抑制されてしまうという問題があった。

【0011】

そこで本発明の目的は、上記問題点に鑑み、イーサネットスイッチが複数接続される環境において輻輳が発生し、PAUSEフレームにより送信トラフィックを制御する際に、そのトラフィック特性を分析することで各トラフィックのQoS、実行スループットの向上、及び各トラフィックの公平性を実現する輻輳制御装置を提供することにある。

【0012】

【課題を解決するための手段】

本発明によれば、イーサネットスイッチにおける輻輳制御装置であって、互いに異なる優先度を有する複数の送信キューと、PAUSEフレームを受信する受信手段と、受信したPAUSEフレームにより前記送信キューからの送信トラフィックを抑制する抑制手段と、を有し、前記抑制手段は、PAUSEタイム以外の時に受信したPAUSEフレームによって、最も優先度の低い送信キューから

の送信トラフィックを抑制し、P A U S E タイム中に受信した P A U S E フレームによって、より優先度の高い送信キューからの送信トラフィックを抑制する輻輳制御装置が提供される。

【 0 0 1 3 】

また本発明によれば、イーサネットスイッチにおける輻輳制御装置であって、送信キューと、P A U S E フレームを受信する受信手段と、受信した P A U S E フレームにより前記送信キューからの送信トラフィックをシェーピングするシェーピング手段と、を有し、前記シェーピング手段は、受信した P A U S E フレームによって、前記送信キューからの送信トラフィックを予め定められたシェーピング値に基づく送信速度以下に抑制する輻輳制御装置が提供される。

【 0 0 1 4 】

さらに本発明によれば、イーサネットスイッチにおける輻輳制御装置であって、送信キューと、前記送信キューに滞留するパケットをその入力ポート対応にカウントすることで、輻輳原因となる入力ポートを特定する特定手段と、特定した入力ポートに結合される他のスイッチに P A U S E フレームを送信する送信手段と、を有する輻輳制御装置が提供される。

【 0 0 1 5 】

【発明の実施の形態】

図 3 は、本発明が適用されるネットワーク構成の一例を示したものである。

本例は、先に示した図 1 と基本的には同じ構成を有しており、例えばスイッチ 1 の受信ポートの 1 つに接続されるスイッチ 2 及び 6 や他の受信ポートに接続されるスイッチ 3、7 及び 8 等はそれぞれが P A U S E フレームマルチキャストの伝達単位を構成している。本例でもスイッチ 1 の送信ポートのリンク α で輻輳が発生している。

【 0 0 1 6 】

図 4 は、本発明による第 1 の実施例を示したものである。また、図 5 には、図 4 の制御フローの一例を示している。

本例では Q o S を確保するための輻輳制御を行なう。まず、バッファの閾値等により輻輳を検出したスイッチ 1 から前段の各スイッチ 2 ~ 5 へ P A U S E フレ

ームが送信される（S101及び102）。

【0017】

PAUSEフレームは、各スイッチ2～5内部のCPU21により検出されて（S103）、スケジューラ22へ渡される。ここで、スケジューラ22は、高、中、低の優先度が付された各送信キュー23～25を制御して優先度に基づく送信トラフィックを制御する。例えば、高優先度にはリアルタイム性の要求される動画や音声データ等が、また低優先度にはFTP等のデータ転送が割当てられる。

【0018】

スケジューラ22は、渡されたPAUSEフレームを判断して、PAUSEフレームをPAUSEタイム中以外の時（通常の送信時）に受信した場合には低優先度のキュー25からの送信を抑制する（S104及び105）。一方、PAUSEタイム中に別のPAUSEフレームを受信した場合には、さらに中優先度のキュー24からの送信を抑制する（S104及び106）。

【0019】

そして、PAUSEタイムのタイマが満了するか、又はPAUSE終了通知（PAUSEフレームのタイマ値＝“0”）を受信すると、送信を抑制しているキューの送信を開始する（S107及び108）。このように、本例によれば優先度を考慮したPAUSE動作が可能となる。その結果、本例の場合にはPAUSEフレームを受信しても高優先度の動画や音声等のトラフィックは中断せず、そのQoSは確保される。

【0020】

図6は、本発明による第2の実施例を示したものである。また、図7には、図4の制御フローの一例を示している。

本例ではシェーピングにより実効スループットの向上を図った輻輳制御を行なう。ここでも、先ずバッファの閾値等により輻輳を検出したスイッチ1から前段の各スイッチ2～5へPAUSEフレームが送信される（S201及び202）。

【0021】

PAUSEフレームは、各スイッチ2～5内部のCPU21により検出されて(S203)、シェーパ31へ渡される。次に、シェーパ31は、送信キュー32に対してその送出速度を本来の物理速度の50パーセントに抑制するシェーピング操作を開始する(S204)。

【0022】

図8及び9には、図6のシェーピング操作の一例を示している。

図8の例では、フレーム長時間からギャップを算出している。すなわち、各送信フレームのフレーム長時間とシェーピングレート値(50%)に基づいたギャップを下記式より算出して生成する。

$$\text{ギャップ値[sec]} = \text{フレーム長時間[sec]} \times (100 - \text{シェーピングレート}[\%]) \div 100$$

【0023】

その結果、送信キュー32からの各送信フレーム毎に同じフレーム時間長のギャップが付加されて、例えば送信速度をその物理速度の50パーセントに抑制するシェーピング操作が行なわれる。また、図9の例では、通常は Δt 時間フレームを送出し続けるところを、その半分の $\Delta/2$ 時間だけに抑制することによって50%のシェーピングを実現している。

【0024】

以降、PAUSEタイムのタイマが満了するか、又はPAUSE終了通知(PAUSEフレームのタイマ値="0")を受信すると、シェーピング操作を終了して通常のデータ送信が開始される(S205及び207)。本例では、スイッチ1において輻輳が発生し得ないように前段の各スイッチ2～5のシェーピング値(%)が予め定められる。

【0025】

すなわち、シェーピング操作時に、前段の各スイッチ2～5の実効送信速度の和が後段のスイッチ1の送信速度を超えないように各スイッチ2～5のシェーピング値が定められる。このように、本例の場合にはPAUSEフレームを受信しても各スイッチ2～5は送信が継続できるため、ネットワーク全体としての実効スループットを向上させる輻輳制御が可能となる。

【 0 0 2 6 】

なお、本例を図 4 の第 1 の実施例と組み合わせて、例えば高優先キュー 2 3 にはシェーピング操作を行なわないとか、また低優先キュー 2 5 の送信を抑制する代わりにシェーピングの程度を 8 0 % に上げる等の処理を行なうことで、Q o S の確保及び実効スループットの向上を両立させる輻輳制御も可能である。

【 0 0 2 7 】

図 1 0 は、本発明による第 3 の実施例を示したものである。また、図 1 1 及び 1 2 には、図 1 0 の制御フローの一例を示している。

本例では、存在する複数のフローに対して公平な帯域の割り当てを行なうことにより、トラフィックの公平性を実現する輻輳制御を行なう。なお、前述した本発明による第 1 及び第 2 の各実施例が前段のスイッチ 2 ～ 5 に適用されるのに対して、本例の場合は次段のスイッチ 1 に適用される。

【 0 0 2 8 】

図 1 0 において、スイッチ 1 は、リンク α への出力ポートの送信キュー 4 2 がある一定の閾値を超えることで輻輳を検出する (S 3 0 1) 。次に、本例では送信キュー 4 2 に滞留するパケットの属性 (例えば、送信アドレスや T C P / U D P のポート番号) が分析される (S 3 0 2) 。この輻輳要因トラフィック分析ルーチンの詳細フローが図 1 2 に示されている。

【 0 0 2 9 】

そこでは、先ず最近の入力フレームの X 個 (例えば、 X = 1 0 0 (個)) のパケットの送信元アドレスがチェックされ、その送信元アドレスをもとに M A C アドレスの学習テーブルやルーティングテーブル等からなるテーブル 4 3 (図 1 0) を参照して対応する入力ポートが検索される (S 4 0 1 及び 4 0 2) 。次に、検索された各入力ポート毎に受信パケット数がカウンタ 4 1 (図 1 0) によってカウントされ、そのカウント値が最大となる入力ポートが特定される (S 4 0 3 及び 4 0 4) 。

【 0 0 3 0 】

図 1 3 は、上記分析処理の一例を図式的に示したものである。

図 1 3 の (a) は、送信キュー 4 2 に滞留するフレーム (パケット) 数が閾値

に達していない場合を示している。やがて、そのフレーム数が図 1 3 の (b) に示すように閾値に達すると、本例では閾値に達したフレームを含めてそれ以前の 1 0 0 フレームの送信元アドレスがチェックされ、テーブル 4 3 を参照しながら対応する入力ポートが検索される。

【 0 0 3 1 】

図 1 0 から明らかなように、記号 ○、△、… はそれぞれ異なる入力ポートからのフレームを現している。従って、前記 1 0 0 フレーム内の各フレーム ○、△、… の数をそれぞれのカウンタ 4 1 によって集計し、例えばフレーム ○ : フレーム △ : フレーム … = 1 個 : m 個 : n 個、ここで $1 > m, n$ であるならば、フレーム ○ の入力ポートが輻輳原因と判断されて輻輳制御の対象となる (S 4 0 3 及び 4 0 4)。

【 0 0 3 2 】

なお、上記の例では最大フレーム数の入力ポートを輻輳原因と判断しているが、他にも例えば全滞留フレームの内の 5 0 % 以上を占めた特定フレームの入力ポートを輻輳原因と判断するようにしてもよい。また、予め設定するフレーム数の幅で常時監視しながら、輻輳が検出された時点でそれまでの滞留フレームの情報から輻輳原因となるフローを識別するようにしてもよい。

【 0 0 3 3 】

図 1 0 及び 1 1 に戻って、スイッチ 1 は輻輳原因として特定したトラフィック (フレーム ○) の入力ポートだけに P A U S E フレームを送出する (S 3 0 3)。図 1 0 の例では、前記入力ポートに接続されるスイッチ 2 に対して P A U S E フレームが送付される。従って、P A U S E フレームを受信したスイッチ 2 だけがこれまでに述べた P A U S E 処理を開始して終了する (S 3 0 4 ~ 3 0 8)。

【 0 0 3 4 】

さらに、P A U S E フレーム内に輻輳原因のトラフィックと特定したパケットの属性 (例えば、送信アドレスや T C P / U D P のポート番号、等) を設定しておき、P A U S E フレームを受信したスイッチ 2 の側でそのトラフィックのみを抑制させるようにすることも可能である。この場合、スイッチ 2 の側に、そのような特定の属性を識別して、対応するトラフィックを動的に抑制する機能を備え

ている必要がある。

【 0 0 3 5 】

このように、本例の場合には輻輳原因となっている特定のスイッチだけが P A U S E 処理の対象となるため、そのスイッチだけがネットワークリソースを占有するような事態は回避され、存在する他のスイッチに対して公平な帯域の割り当てが可能となる。

【 0 0 3 6 】

なお、本例は輻輳を検出する側のスイッチ 1 に関するものであり、その前段のスイッチ 2 ～ 5 に関する第 1 及び第 2 の実施例と組み合わせることにより、Q o S の確保、実効スループットの向上、及びトラフィックの公平性を同時に実現する輻輳制御も可能である。

【 0 0 3 7 】

【発明の効果】

以上説明したように、本発明によれば Q o S を考慮した輻輳制御が可能となる。また、本発明によれば輻輳状態と非輻輳状態の反復状況が回避され、輻輳が発生したネットワークでの実効的なスループットを向上させることができる。さらに、本発明によれば輻輳発生時に輻輳原因となる特定の送信トラフィックのみが抑制され、ネットワークリソースの公平な利用が可能となる。

【図面の簡単な説明】

【図 1】

P A U S E フレームによる従来の輻輳制御の一例を示した図である。

【図 2】

P A U S E フレームのフォーマットの一例を示した図である。

【図 3】

本発明が適用されるネットワーク構成の一例を示した図である。

【図 4】

本発明による第 1 の実施例を示した図である。

【図 5】

図 4 の制御フローの一例を示した図である。

【図 6】

本発明による第 2 の実施例を示した図である。

【図 7】

図 6 の制御フローの一例を示した図である。

【図 8】

図 6 のシェーピング操作の一例 (1) を示した図である。

【図 9】

図 6 のシェーピング操作の一例 (2) を示した図である。

【図 1 0】

本発明による第 3 の実施例を示した図である。

【図 1 1】

図 1 0 の制御フローの一例 (1) を示した図である。

【図 1 2】

図 1 0 の制御フローの一例 (2) を示した図である。

【図 1 3】

トラフィック分析処理の一例を図式的に示した図である。

【符号の説明】

1 ~ 1 1 ... スイッチ

2 1 ... CPU

2 2 ... スケジューラ

2 3 ~ 2 5、3 2、4 2 ... 送信キュー

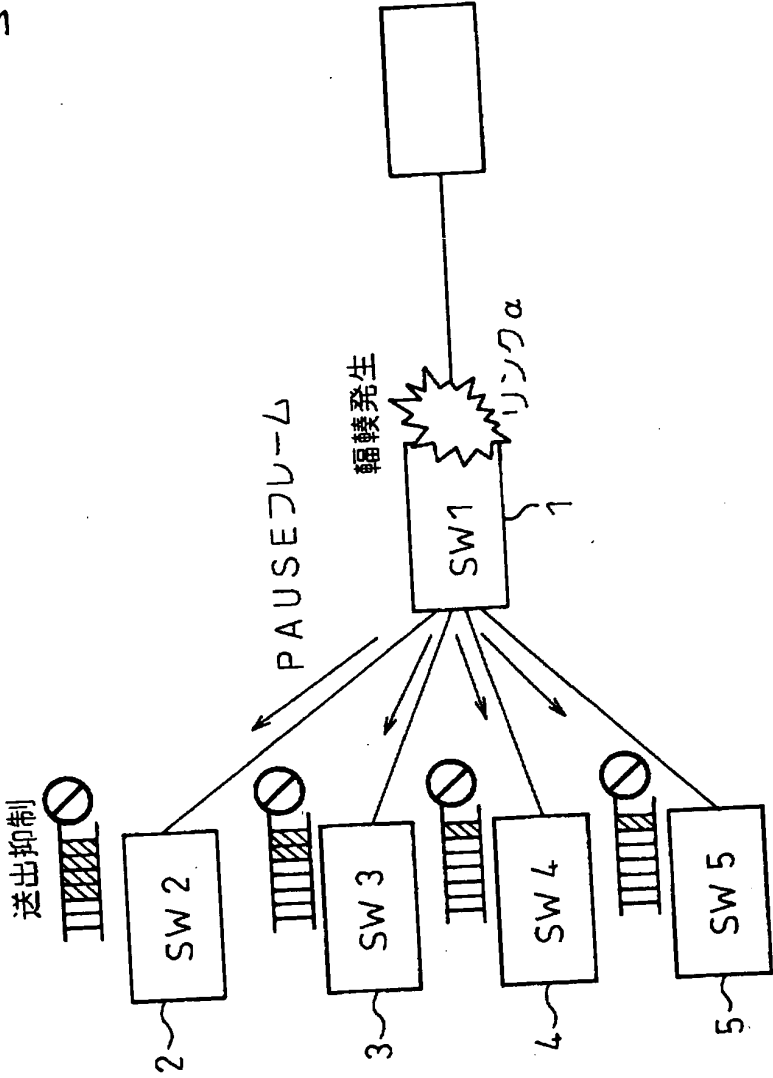
3 1 ... シェーパ

4 1 ... カウンタ

4 3 ... テーブル

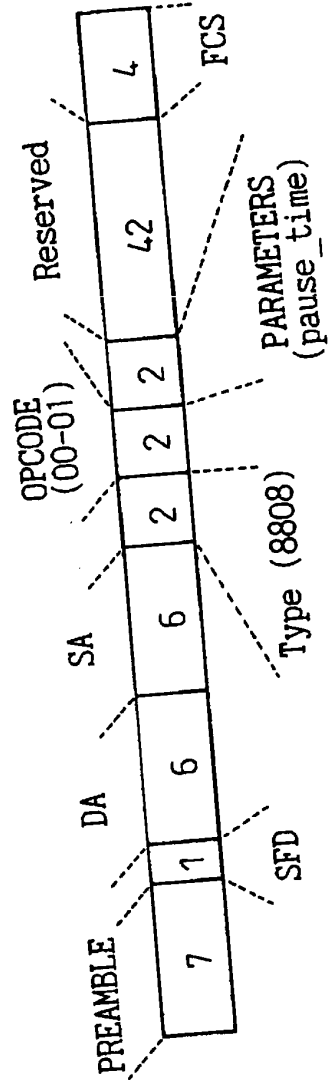
【書類名】 図面
【図1】

図 1



【図2】

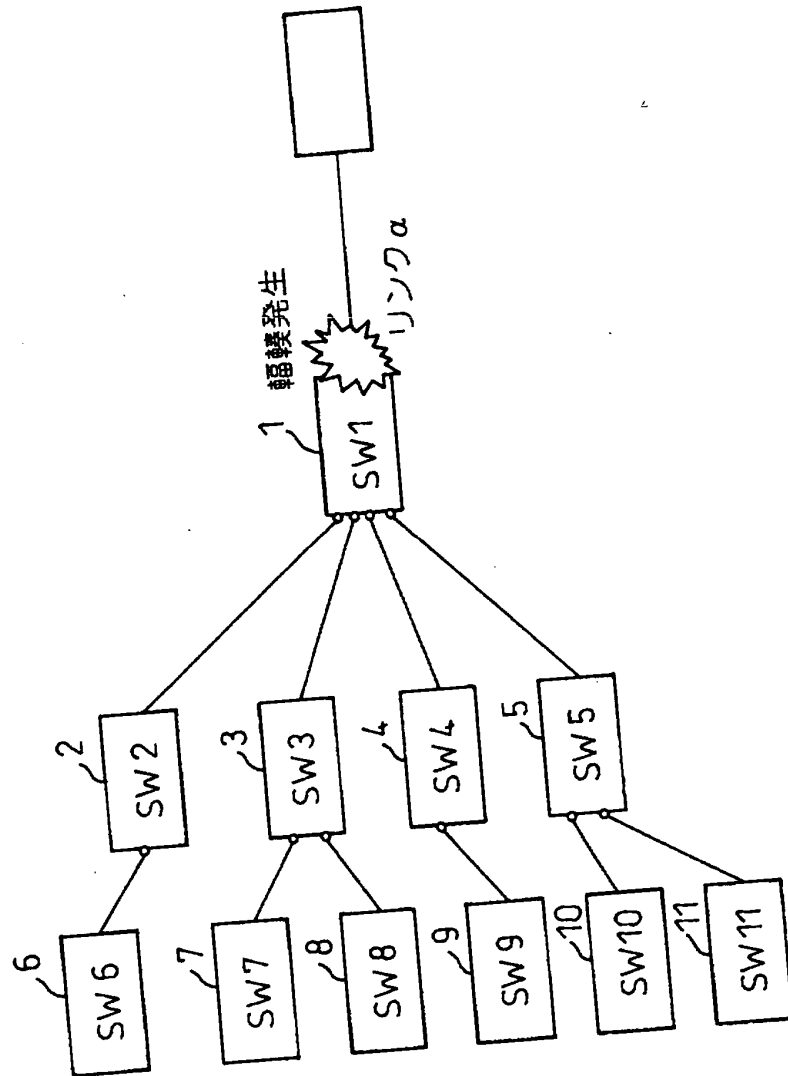
図2



Type: 0x8808
 DA: 01:80:C2:00:00:01 (multicast)
 OPc: 0x0001
 Parameter: 0~65535

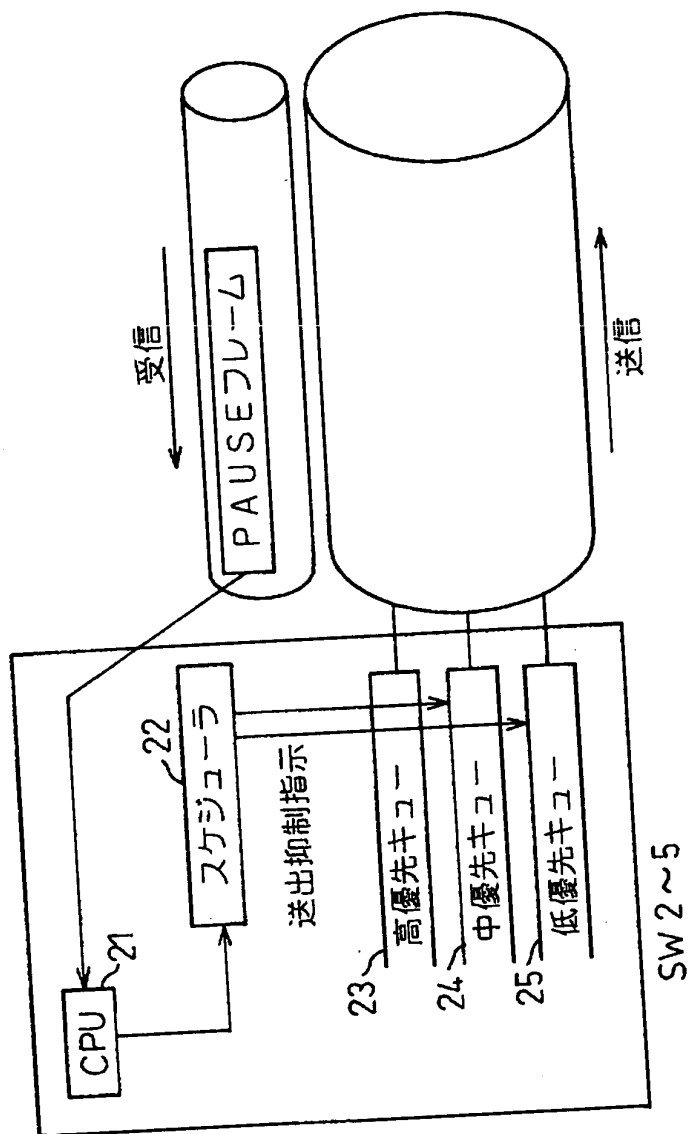
【図3】

図3



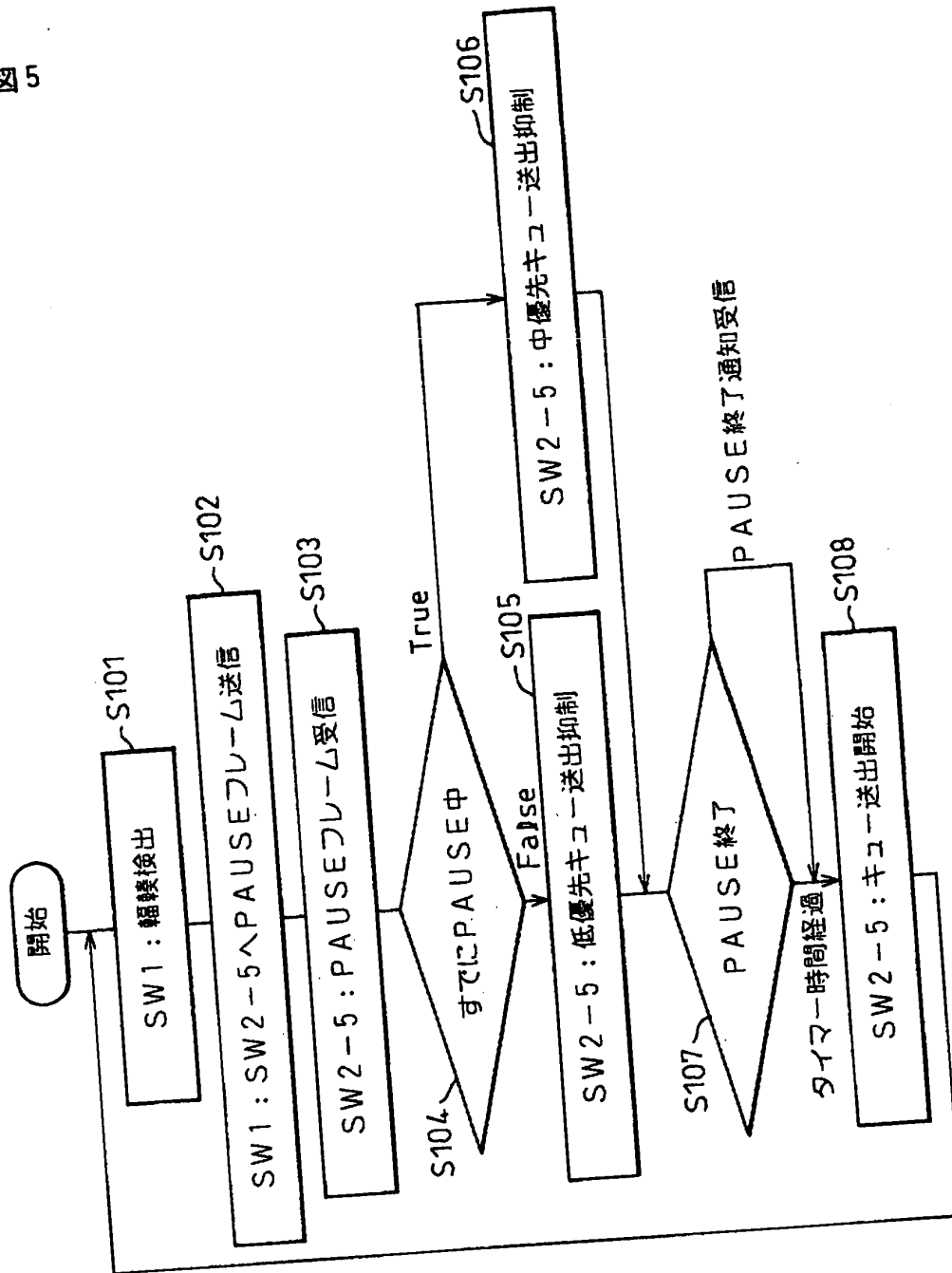
【図4】

図4



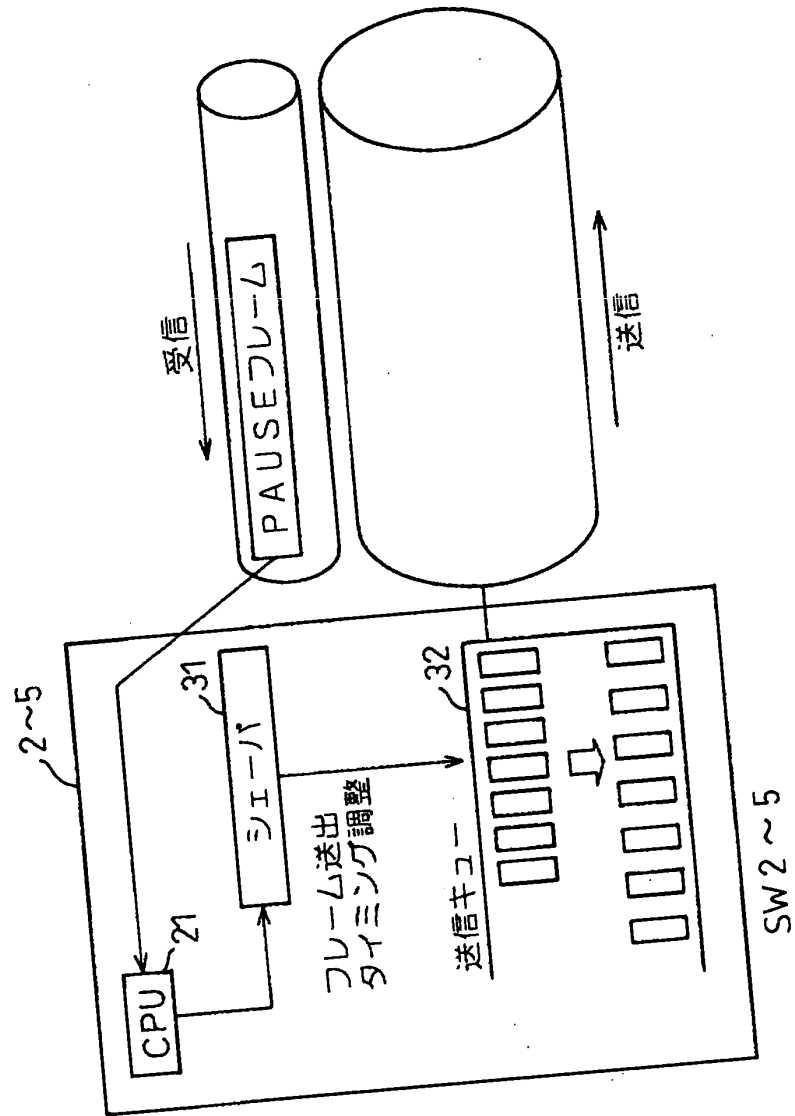
【図5】

図5

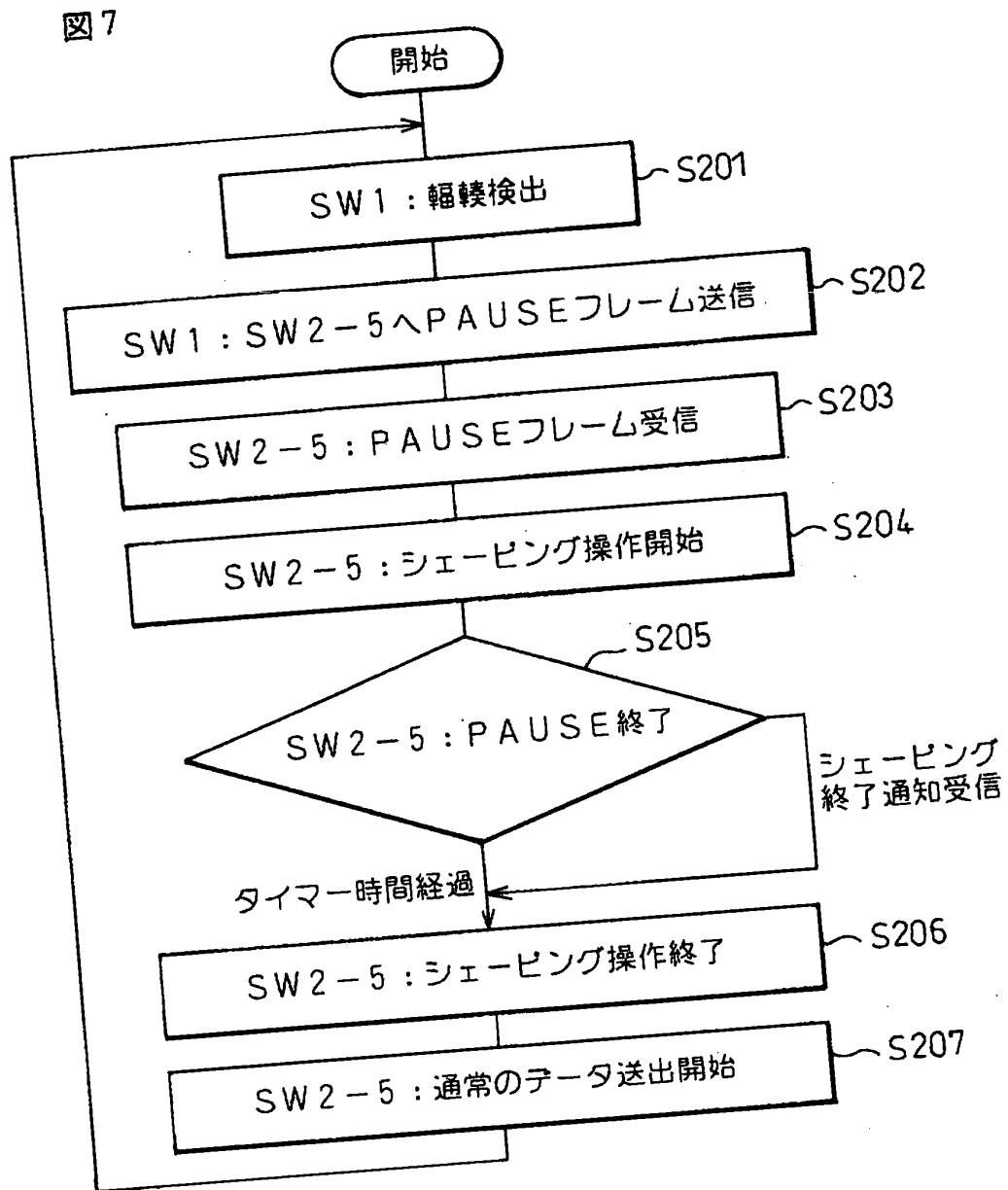


【図6】

図6



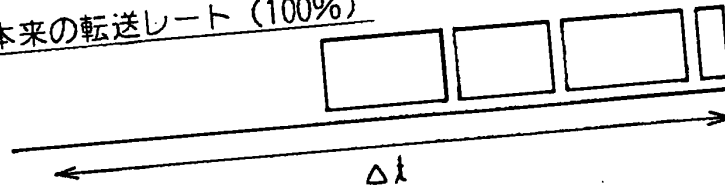
【図7】



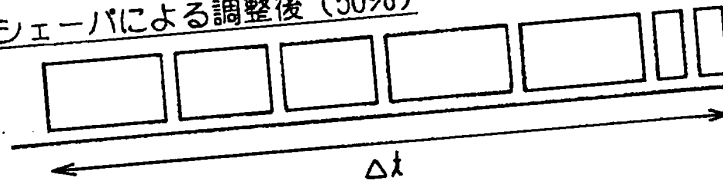
【図8】

図8

方式Ⅰ フレーム長時間からギャップを算出
本来の転送レート (100%)



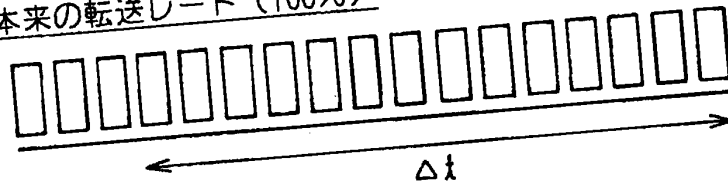
シェーパによる調整後 (50%)



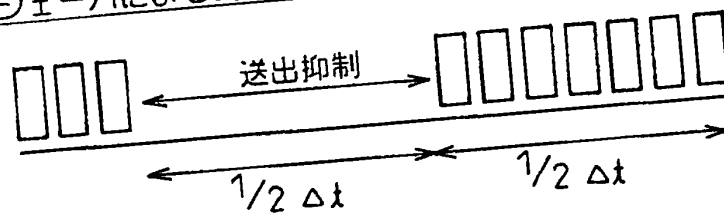
【図9】

図9

方式Ⅱ Δt 時間でのギャップ生成
本来の転送レート (100%)

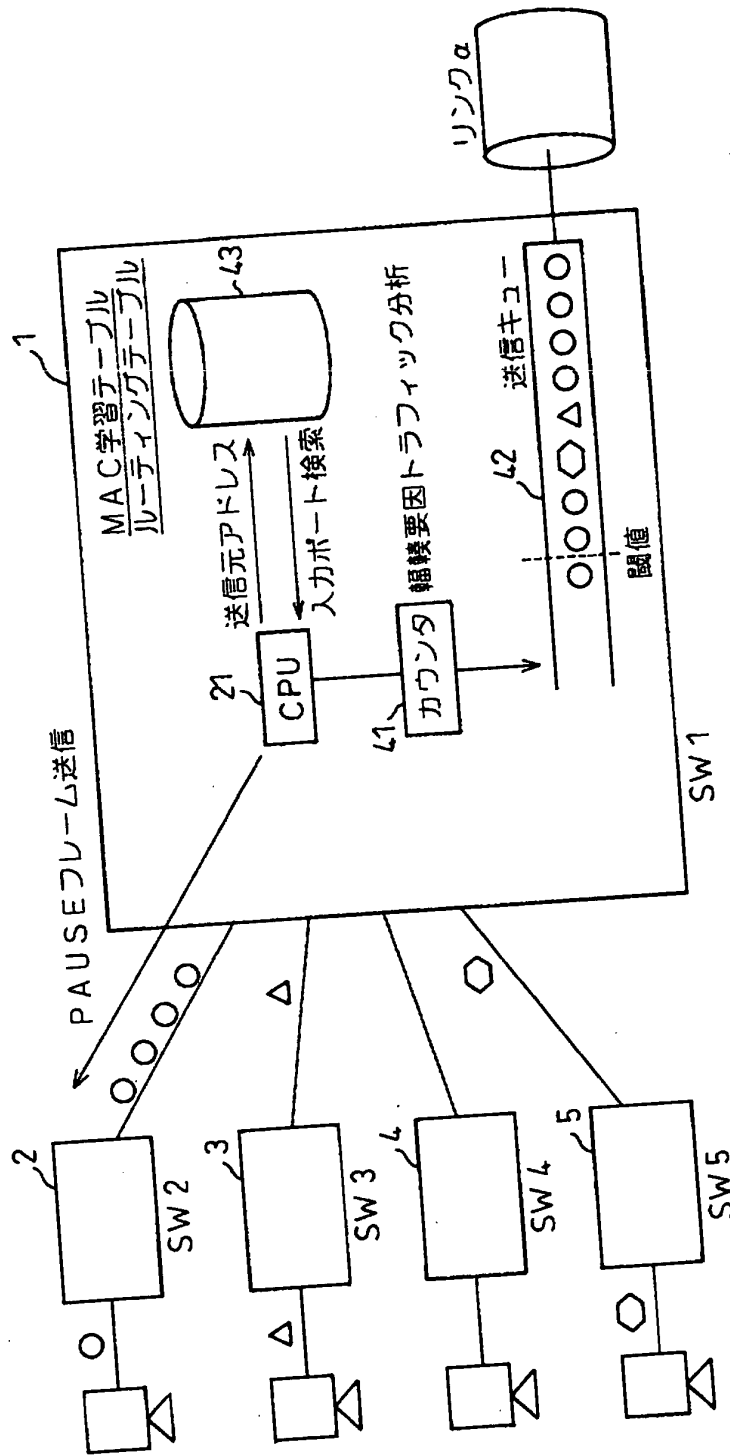


シェーパによる調整後 (50%)



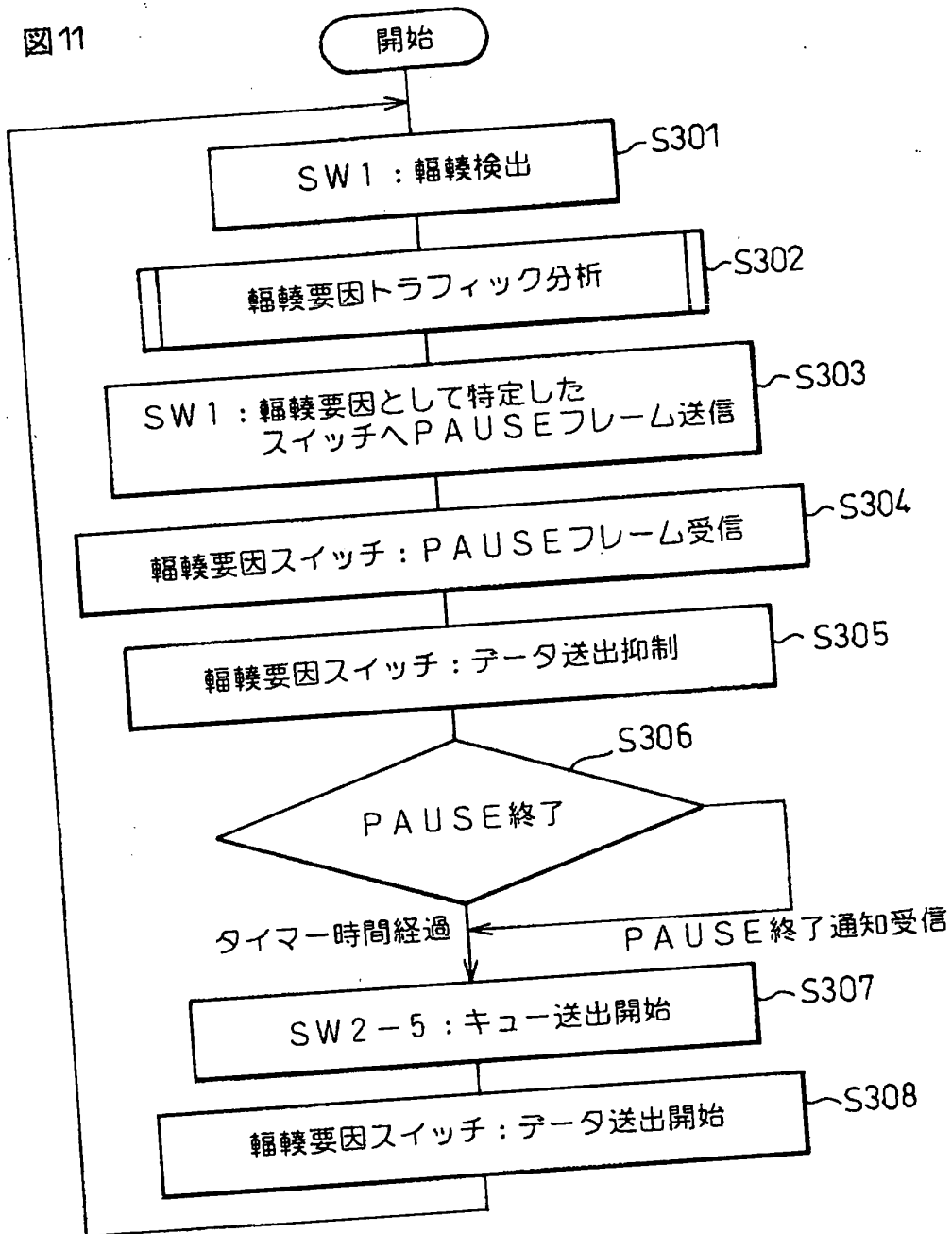
【図10】

図10

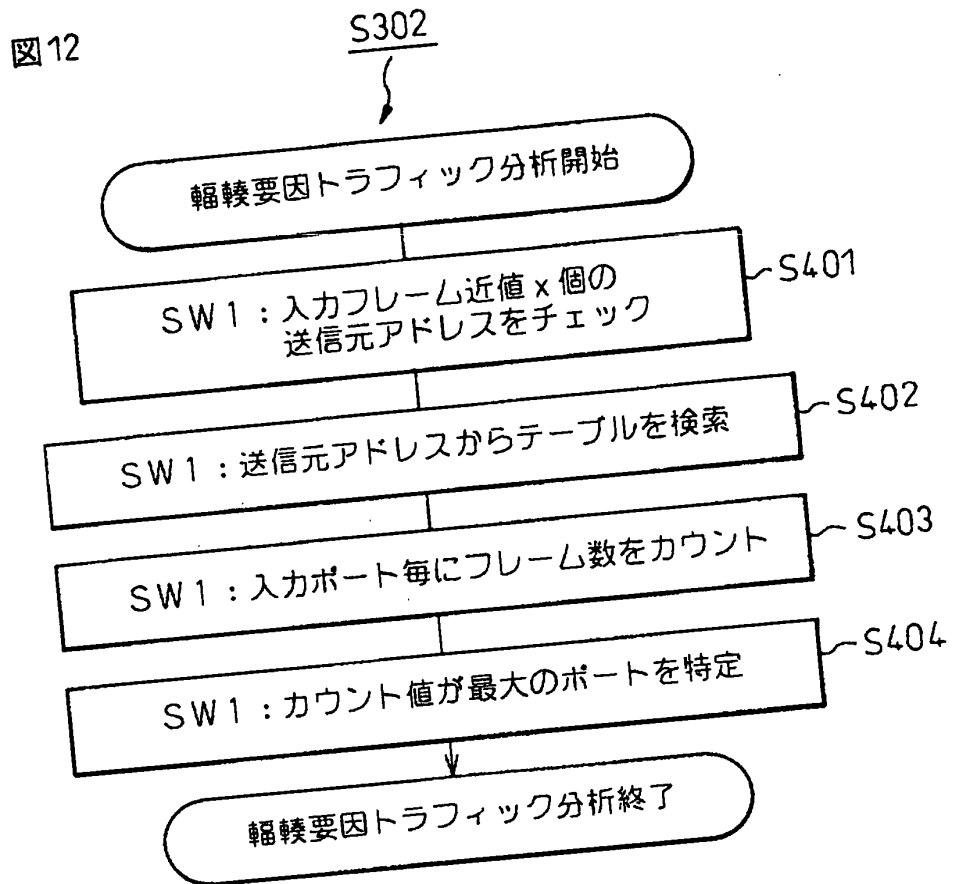


【図11】

図11

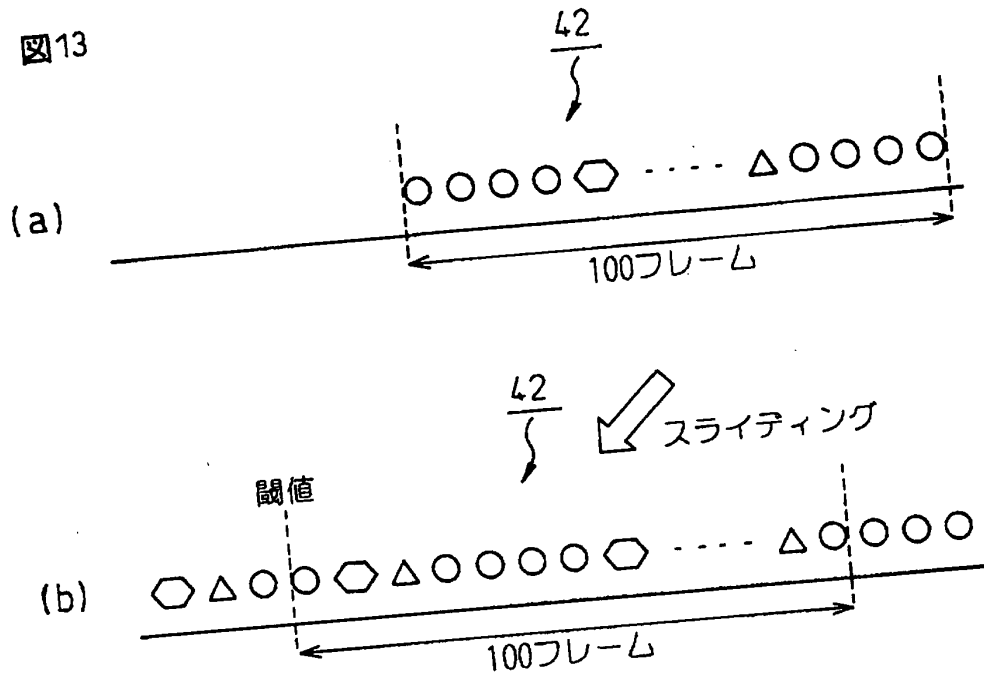


【図12】



【図13】

図13



【書類名】 要約書

【要約】

【課題】 輻輳制御装置に関し、特にイーサネットスイッチが複数接続される環境において輻輳が発生すると、トラフィック特性に基づいて送信トラフィックを抑制する輻輳制御装置を提供する。

【解決手段】 イーサネットスイッチにおける輻輳制御装置は、互いに異なる優先度を有する複数の送信キューと、PAUSEフレームを受信する受信手段と、受信したPAUSEフレームにより前記送信キューからの送信トラフィックを抑制する抑制手段と、を有し、前記抑制手段は、PAUSEタイム以外の時に受信したPAUSEフレームによって、最も優先度の低い送信キューからの送信トラフィックを抑制し、PAUSEタイム中に受信したPAUSEフレームによって、より優先度の高い送信キューからの送信トラフィックを抑制する、ことから成る。

【選択図】 図4

特2002-263102

出願人履歴情報

識別番号

[000005223]

1. 変更年月日

1996年 3月26日

[変更理由]

住所変更

住所

神奈川県川崎市中原区上小田中4丁目1番1号

氏名

富士通株式会社